Received: 7 December 2020 | Revised: 9 August 2021 | Accepted: 20 August 2021

DOI: 10.1111/desc.13174

PAPER

Developmental Science WILEY

Post-error slowing: Large scale study in an online learning environment for practising mathematics and language

Susanne M. M. de Mooij[1] | Iroise Dumontheil[1,2] | Natasha Z. Kirkham[1] | Maartje E. J. Raijmakers[3] | Han L. J. van der Maas[3]

[1] Department of Psychological Sciences, Centre for Brain and Cognitive Development, Birkbeck, University of London, London, UK

[2] Centre for Educational Neuroscience, University of London, London, UK

[3] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

Correspondence
Susanne M.M. de Mooij, Department of Psychological Sciences, Centre for Brain and Cognitive Development, Birkbeck, University of London, Malet street, London, WC1E 7HX, UK.
Email: sdemoo01@mail.bbk.ac.uk

Online learning platform: Data was collected in the online learning environment of Prowise Learn (www.oefenweb.com), which originates from the University of Amsterdam in the Netherlands.

Funding information
H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 721895

Abstract

The ability to monitor and adjust our performance is crucial for adaptive behaviour, a key component of human cognitive control. One widely studied metric of this behaviour is post-error slowing (PES), the finding that humans tend to slow down their performance after making an error. This study is a first attempt at generalizing the effect of PES to an online adaptive learning environment where children practise mathematics and language skills. This population was of particular interest since the major development of error processing occurs during childhood. Eight million response patterns were collected from 150,000 users aged 5 to 13 years old for 6 months, across 23 different learning activities. PES could be observed in most learning activities and greater PES was associated with greater post-error accuracy. PES also varied as a function of several variables. At the task level, PES was greater when there was less time pressure, when errors were slower, and in learning activities focusing on mathematical rather than language skills. At the individual level, students who chose the most difficult level to practise and had higher skill ability also showed greater PES. Finally, non-linear developmental differences in error processing were found, where the PES magnitude increased from 6 to 9-years-old and decreased from 9 to 13. This study shows that PES underlies adaptive behaviour in an educational context for primary school students.

KEYWORDS
adaptive behaviour, educational games, online learning environment, post-error slowing

1 | INTRODUCTION

Making mistakes is part of the learning process, allowing a student to learn and improve (Moser et al., 2011). The capacity to detect, evaluate and adapt to these mistakes is essential to the development of children and is an active area of research in the cognitive control literature (Regev & Meiran, 2014; Smulders et al., 2016; Tamnes et al., 2013; Ullsperger et al., 2014). One of the studied markers of adaptive behaviour is post-error slowing (PES). PES refers to the finding that humans slow down their performance after making an error, such that the reaction time (RT) after an error is greater than after a correct response (Laming, 1979; Rabbitt & Rodgers, 1977). Whilst prior studies have typically examined PES within experimental settings using executive function tasks, the present study examined PES outside the laboratory in a range of tasks implemented in an online learning environment designed to allow children to practise mathematics and language skills. This allowed the investigation of individual and task factors that affect PES.

RESEARCH HIGHLIGHTS

- Children slow down their performance after an error (PES), an overall marker of cognitive control, while practising in an online learning environment.
- Greater slowing down is associated with improved performance after an error, suggesting that PES is adaptive.
- The occurrence of adaptive behaviour differed between mathematical and language tasks and depended on the amount of time pressure.
- Children's ability and the level of difficulty at which they chose to practise predicted the occurrence of PES.
- Development differences in PES showed a non-linear pattern, with a peak at age 9, suggesting increasing awareness of errors and ability to change behaviour.

1.1 | Theoretical accounts of PES

The Dual Mechanisms of Control (DMC) framework proposes that cognitive control comprises reactive and proactive components which can be used adaptively by adults depending on the demands of a task (Braver, 2012). With proactive control, individuals anticipate and prepare to respond to upcoming events. This approach can reduce detrimental interference during the events but is quite demanding on working memory capacity. In contrast, reactive control allows individuals to respond to unforeseen events, such as an error. The reactive mode is less demanding on working memory but also less efficient and takes time and may therefore lead to PES. Two main theoretical accounts of PES have been proposed. The functional account argues that PES serves the purpose of taking more time to plan an action to prevent future errors and increase performance. This is supported by the conflict monitoring theory (Botvinick et al., 2001), which states that slowing down reflects increased reactive control evoked by a conflict, in this case, an error. People monitor their performance and consequently adjust their response thresholds leading to slower but more accurate responses. This was also shown by Dutilh et al. (2012b)'s drift-diffusion model, where participants increased their boundary separation, the model parameter for response caution, after making an erroneous decision. A more cautious response leading to more accurate behaviour in post-error trials is thought to underly adaptive behaviour (Ridderinkhof et al., 2004), although a coupling of post-error changes in accuracy (PEA) and reaction time is not always found (for a review, see Ullsperger et al., 2014). An alternative explanation supporting the functional account is considering PES as the product of automatic inhibition of the next response after an error (Gupta et al., 2009; Marco-Pallarés et al., 2008), where inhibitory mechanisms increase response cautiousness after making an error.

Other studies have suggested a non-functional account of PES, where an error is thought to have a negative effect on subsequent post-error performance. The orienting account argues that PES occurs especially with infrequent errors since the attention orients towards this surprising event instead of the task (Houtman et al., 2012; Notebaert et al., 2009). This attention shift disrupts the information process, inducing slower responses and worse performance. Notably, Danielmeier and Ullsperger (2011) point out in their review that there is evidence for both the functional and non-functional account and that they are not necessarily mutually exclusive, as multiple mechanisms may contribute to PES.

Although PES effects have been presented as markers of adaptive behaviour, the majority of studies are done with relatively simple and rapid tasks such as the flanker task (e.g. Schroder et al., 2019), or the Go/No-go task (e.g. Jonker et al., 2013). These tasks are limited in terms of the adaptation strategies they allow and performance can only be increased by paying more attention to the stimuli, in line with Notebaert's (2009) conflict monitoring theory. However, there are studies that have investigated PES in more complex settings with adults, for example using a grasping task (Ceccarini & Castiello, 2018). In academic context, studies have also used mental arithmetic tasks with university students and found that multiple strategies were available, with larger response times after an error than after a correct response (Borght et al., 2016; Desmet et al., 2012; Lavro et al., 2018b; Núñez-peña et al., 2017). Interestingly, Borght et al. (2016) found that switching to a different strategy was associated with an increase in PEA and reduced PES. Recently, children aged 4 through 15 were tested outside the laboratory in their Montessori classrooms using a flanker task and showed post error slowing with increased self-monitoring performance later (Denervaud et al., 2020).

1.2 | Development of PES

Depending on the context, adults flexibly switch to the most adaptive mode of control, either reactive or proactive (Chiew & Braver, 2013). Cognitive control is an ability that children are still developing: younger children seem to mostly rely on the reactive control mode (even when the proactive control is more efficient), but begin to engage in proactive control from the age of 8 years (Chevalier et al., 2015; Niebaum et al., 2020). The development of cognitive control is not only associated with the improvement of these core cognitive processes, but also with improvements in the adaptive selection of a mode of control to engage in in a specific context or point in time. The development of error monitoring as measured by PES has mostly been examined using neuroscientific methods with children between 5 and 12 years of age (for reviews, see Ferdinand & Kray, 2014; Tamnes et al., 2013). Some behavioural studies such as Fairweather (1978) found that young children from the age of 5 slowed down after making errors during a two- to eight-choice RT task. Ever since, PES in children has been repeatedly examined and observed (Berwid et al., 2014; Fairweather, 1978; Gupta et al., 2009; Jones et al., 2003; Schachar et al., 2004; Smulders et al., 2016). Following the orienting account where errors lead to interference, children are thought to be more prone to interference than young adults and therefore also exhibit more PES (Smulders et al., 2016; van der Molen, 2000). Conflict arising from interference is also related to children's still-developing inhibitory control skills (Durston et al., 2002; Garon, Bryson, & Smith, 2008; Welsh, Friedman, & Spieker, 2006; B. R. Williams, Ponesse, Schachar, Logan, & Tannock, 1999). This indirect relation implies that improving inhibitory control skills would therefore lead to a reduction in PES magnitude. On the other hand, greater PES may also reflect a developmental increase in cognitive control, with greater performance monitoring and strategic adjustment of the balance between proactive and reactive control, according to the functional account (Jones et al., 2003; Tamnes et al., 2013).

The directionality of changes in PES during development described in the literature is heterogeneous. In recent research, Smulders et al. (2016) used standard two-choice RT tasks and found that post-error response slowing was present from 5 years to adulthood, i.e., the whole of their examined age range. They reported PES stability with age rather than developmental differences. Jones et al. (2003) reported increasing PES from 3- to 4-years of age using a Simon Says inhibitory control task. These changes were interpreted as reflecting a developmental increase in cognitive control. The results of Gupta et al. (2009) showed a non-linear development of PES in 6-11-year-old children performing two task-switching digit tasks, with a peak in magnitude of PES at age 7. Schachar et al. (2004) also found a decreasing PES magnitude from 7 to 16 years of age in a stop-signal task.

1.3 | Factors influencing the presence or magnitude of PES

Prior research has shown that PES varies as a function of the type of task as well as the difficulty of the task. Although PES has been found in both easy tasks (e.g., flanker task, Schroder et al., 2020) and more complex tasks (e.g., mental arithmetic task, Desmet et al., 2012), the non-functional account noted that response slowing after an error occurs mostly under conditions when errors are rather unexpected and infrequent events (Danielmeier & Ullsperger, 2011; Lavro et al., 2018a; Notebaert et al., 2009).

Another factor influencing PES is the type of error that preceded it. Errors can be found to be systematically faster or slower than correct responses, with a range of underlying causes (Ratcliff & Rouder, 1998). Slow errors typically occur when accuracy is emphasized and the task is relatively difficult, whereas fast errors occur when responding is rushed. Damaso et al. (2020) categorised fast and slow responses as the 50% slowest and fastest responses of each participant in two simple recognition memory experiments. Results showed that PES mostly occurred after fast errors (Damaso et al., 2020).

1.4 | The current study

In the present study, we investigated students' post-error performance in a large scale online adaptive learning environment for primary school children aged 5–13 years old. The first aim of this study was to assess the presence of PES in the various learning activities relating to different mathematical and language skills. In addition, we examined whether PES associated with increased PEA, which was expected if PES functionality is adaptive of nature, but not if PES reflected interference, as proposed in the orienting account.

The second aim was to investigate predictors of the magnitude of PES and of the association between PES and PEA. At the task level, we investigated whether the type of skill practised (mathematics vs. language) was associated with the magnitude of PES. In the learning environment used for this study, children can choose not only between different types of games but also the difficulty level of the games (hard, medium and easy), which relate to the probability of solving the items correctly based on their ability (Brinkhuis et al., 2018; Jansen et al., 2013). This allowed us to test the prediction, based on previous research (Danielmeier & Ullsperger, 2011; Lavro et al., 2018a; Notebaert et al., 2009), that children who choose the easy level, where there is less chance of making errors, were more likely to show PES after an error than children playing the medium and hard level, as the errors were unexpected.

Finally, we investigated the influence of the speed of responding on the PES. We first assessed the effect of time pressure (i.e., how much time children were given to respond in each task) on PES magnitude. Secondly, we distinguished fast and slow errors by categorising whether RT was higher or lower than the median split of the correct trials before the error within a task (Damaso et al., 2020) – to control for, global fluctuations in skills and motivation of the participant. The prediction was that trials with fast errors would lead to greater PES.

We also assessed individual differences between participants as potential predictors of PES as well as of the PES-PEA association. We first examined associations with age. Given the development of cognitive control, it was expected that younger children would show more PES, reflecting greater reliance on reactive control, which is not necessarily beneficial for accuracy. Older children were expected to show a greater PES-PEA association, reflecting their improving proactive control skills. Second, we investigated how individual differences in PES may associate with children's ability in the task examined. This has not been investigated before, however, large differences in mathematical abilities occur within school grades (Straatemeier, 2014), so only considering the age of children may not be a good proxy for comparing children's developmental stage. Therefore, we predicted that ability on a particular task would be associated with PES and PEA above and beyond age.

2 | METHODS

2.1 | Participants

The response data of 149,747 Dutch primary school children playing in the learning environment were collected. Their age was between 5 and 13 years old ($M = 9.4$ y, $SD = 1.8$ y, 48.6% female). Primary schools buy accounts for students to practise their language and mathematical skills in the learning environment, while their responses are logged for scientific purposes such as this study. Children (their parents or schools) can opt out of being part of the research done in the learning environment, in which case they were not included in this study. All anonymized data are available to researchers, and access to the data can be acquired by contacting the first author.

2.2 | Materials and equipment

2.2.1 | Learning environment

Data were collected in the online adaptive learning platform *Prowise Learn* (www.oefenweb.nl) with games to practise mathematics and language skills, actively used by Dutch primary school children. The adaptivity of this system is determined by an on-the-fly Elo-based estimation algorithm based on the accuracy and speed of the students. This approach is named after Arpad Elo, who originally developed it for chess competition ranking. Here, a person's ability rating increases when they solve the problem correctly and fast and decreases when the answer is incorrect or very slow, and vice versa for the item difficulty ratings (for more detail, see Klinkenberg, Straatemeier, & Van Der Maas, 2011; Maris & van der Maas, 2012). Based on a student's current ability estimate, the difficulty level of the items presented to the students can be set so that the student has a probability of .90 (easy level), .75 (medium level), or .60 (hard level) of answering correctly, to ensure students remain motivated (Jansen et al., 2013; Straatemeier, 2014). This set difficulty level can be chosen by the student and changed when preferred (Brinkhuis et al., 2018; Wilson et al., 2019).

2.2.2 | Arithmetic and language learning activities

Every student has his/her own virtual garden, where each plant represents a game to practise a specific ability. In this study, 23 different learning activities (i.e., games) were analysed. Each game session consists of ten problems where a response must be given within a certain time limit, which is visualised as virtual coins counting down on the bottom right of the screen. The student receives immediate feedback on the accuracy of their response; the remaining coins are rewarded after a correct response and subtracted when the response is incorrect. No coin is earned or lost when the student fails to answer within the time limit. This way of scoring is known as the 'High Speed High Stakes' rule (for more details, see de Mooij et al., 2020; Maris & van der Maas, 2012).

2.3 | Procedure

Participants decide for themselves when they play, how long they play for and which learning activities they want to complete. Data from a total of 45 million trials were collected in 6 months (June 2019 until November 2019) from 23 different learning activities (13 mathematics-related and 10 language-related, see Appendix A.1 for more details about the learning activities). Since the students decide themselves which learning activity they participate in and these learning activities vary in which minimum age is required to practise, the number of participants and the average age is different for each activity (see Appendix A.1 for details). Trials with an RT faster than 200 ms were regarded as guess responses and were therefore excluded. In the learning environment, children are discouraged to guess due to a visible and direct penalty of fast and incorrect responses. Trials where no response was made within the time limit were labelled as missing.

To ensure a reliable measure of post-error behaviour, RTs were selected from a specific pattern of response. In a game session, a sequence of four problems was selected when the accuracy pattern was 1-1-0-1 (1 = correct; 0 = incorrect), such that a minimum of two correct responses precede an error and the trial after the error is correct. The additional correct pre-error response was added to the response sequences to ensure that a correct post-error response in a session could not at the same time be a correct pre-error response. In the 45 million trials, 8 million of such sequences were found. Similarly for the PEA measure, all occurrences of the pattern 1-1-0, i.e., a minimum of two correct responses followed by an error, were identified in the same dataset. Next, a proportion PEA was computed based on the accuracy of the first trial following this pattern for each learning activity and each participant.

2.4 | Measures

2.4.1 | Post-error slowing quantification

The majority of previous studies quantify the magnitude of PES as the difference in mean RT between trials following an error and trials following a correct response. As Dutilh et al. (2012b) pointed out, this method can be confounded by the global fluctuations in ability and motivation during the task, since post-error responses are more likely to originate from the second half of a task where responses are inevitably slower due to motivation and tiredness. Dutilh's solution is to quantify PES as the average, across the selected sequences, of the difference between the RT after the error (RT_{E+1}) and the RT before the error (RT_{E-1}) (PES_{diff}):

$$\overline{PES_{diff}} = MRT_{E+1} - MRT_{E-1} \tag{1}$$

To ensure that our results do not fully depend on the choice of this absolute PES measure, we use two additional methods. We also report PES relative to the overall ability speed, calculated by dividing the RT difference with the average speed of the two trials (i.e., $\frac{1}{2}$ ($MRT_{E+1} + MRT_{E-1}$)). This ensures that individual differences in PES magnitude can be compared regardless of the overall speed/ability of the student:

$$\widehat{PES_{rel}} = \frac{MRT_{E+1} - MRT_{E-1}}{\frac{1}{2}\ (MRT_{E+1} + MRT_{E-1})} \tag{2}$$

The third method is a robust way of measuring post-error behaviour overall, by quantifying PES as the number of sequences where the RT was larger after the error (RT_{E+1}) than before the error (RT_{E-1}), relative to the number of sequences (N). This is a measure that is not affected by the PES effect size in certain sequences, which can vary considerably in such a big dataset. The disadvantage of this robust method is its low power.

$$\widehat{PES_{robust}} = \frac{n\,(RT_{E+1} > RT_{E-1})}{N}$$

2.5 | Analyses

The main analysis focused on whether PES could be found for different learning activities. To do this, a linear mixed model was performed in R (Team, 2013) using the package *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) to calculate *p*-values. Participant was treated as a random effect because the number of sequences collected per participant differed as participants can play whenever they want and as much as they want. Learning activity was treated as a fixed effect to investigate whether there were differences in PES effect per learning activity. When a general PES effect was found, the PEA measure was added along with a variety of predictors to a basic linear mixed model where both learning activity and participant were treated as random effects, to account for variance between the tasks and participants. A chi-square test was used to see if the added predictors explain PES better.

The first categorical predictor added, was the *type of learning activity*, distinguishing between mathematical and language skills. The second task predictor was *time pressure* since the tasks differed in the time limit given to participants to solve a problem, from 8 to 60 s (see Appendix A.1). The third categorical predictor, *difficulty level* (easy, medium, hard level), reflected the probability of answering correctly chosen by the child. At the participant level, we also included age (5-13 years old) and child's ability level (a continuous scale from −10 to 10) as predictors of PES magnitude. These last two predictors were analysed in a linear and quadratic fashion since the study of Gupta et al. (2009) also found a non-linear development of PES with age. The ability level of a child in each learning activity is provided by the learning system itself. It is estimated with an adaptive Elo algorithm and is updated after every item. Lastly, *type of error* was investigated at the trial level, where a comparison of the magnitude of PES was made between fast and slow errors. All errors were categorised as *fast error* when the RT was lower than the median RT of the correct items before the error or as *slow error* when the RT was higher than the median RT.

3 | RESULTS

3.1 | Overall PES effect

To measure whether there was PES in the learning activities, a linear mixed model was fitted to the data with learning activity as fixed effect and participant as random effects. Using the absolute PES_{diff} measure we found that in 20 of the 23 learning activities participants showed a significantly larger RT after an error than before, with a mean difference across learning activities of 225 ms (Figure 1 and Appendix A.2 for details on the statistics). The average age of participants differed considerably between the learning activities (8.2 – 10.8 years), but this did not account for the variation in PES magnitude between the learning activities, $r = -.08$, $p = .70$. For the PES_{rel} measure we found the same pattern, where the learning activities *letterchaos* and *spelling* did not show PES but post-error speeding (Appendix A.2). The learning activity practising grammar was not significantly different from zero in PES_{diff}.

Since PES_{robust} is a proportion measure, we performed a separate proportion test for every learning activity, testing whether RTs were greater after an error than before on more than 50% of trials, corresponding to a mean proportion score greater than 0.5. The results were in line with the linear mixed models. The average proportion across learning activities was .52 [range .48 – .55], meaning that in 52% of the sequences the RT after an error was greater than before the error.

3.2 | PES-PEA association

PES estimates per learning activity were expected to be positively associated with PEA, reflecting an adaptive response. This association was analysed with a one-sided Pearson correlation test. Only the learning activities with a PES_{diff} greater than zero were analysed since these showed PES. The test indicated a positive PES-PEA association, $r = .437$, $t(19) = 1.72$, $p = .049$, such that the learning activities showing greater PES also showed greater post-error accuracy (Figure 2A). A linear mixed model of PES data at the trial-level with PEA per learning activity and participant as random effect predictors also predicted PES magnitude, $\beta = 96.90$, $t(121000) = 2.36$, $p = .018$.

3.3 | Type of learning activity and time pressure

Next, a variety of predictors were added and compared to a basic model with participant and learning activity as random effects to see whether these predictors could explain variance in the magnitude of PES. Learning activity PEA was also included as a fixed effect in all the analyses, to look at whether the predictors influenced the PES-PEA association. Since the different PES calculations showed comparable results, we only discuss PES_{diff}.

We first investigated task level predictors of PES, to understand why the learning activities differed in PES magnitude. We found that type

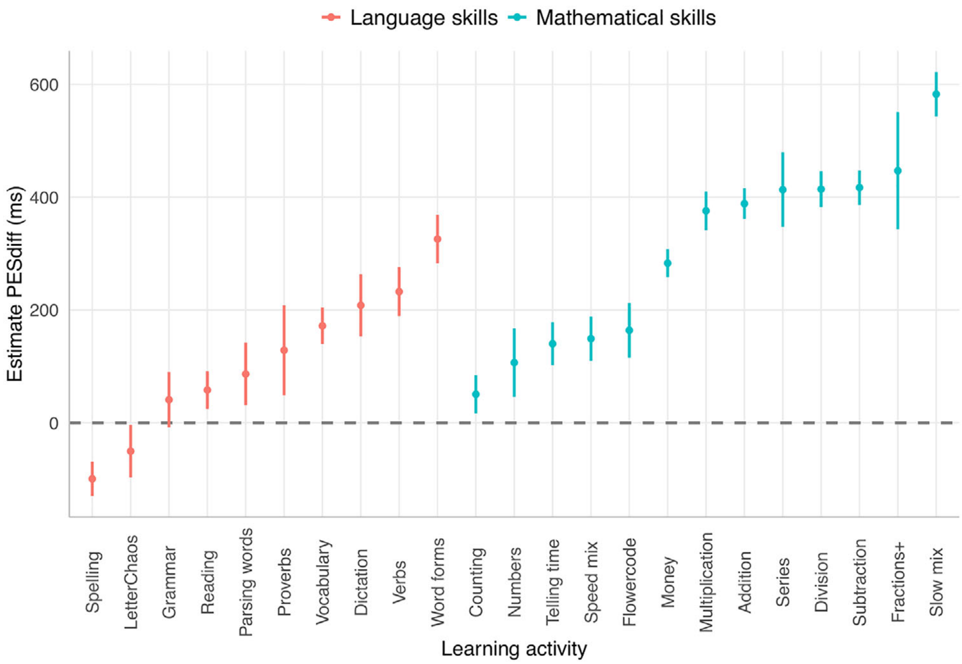

FIGURE 1 Estimated absolute post-error slowing (PES_{diff}) in the 23 learning activities divided into practising language skills (red) and mathematical skills (blue). PES was observed for all activities (coefficients above zero (dashed line)), except for letterchaos and spelling (see Appendix A.1 for the game details). The points are the regression coefficients estimated from the linear mixed model; the vertical lines represent the 95% confidence intervals

of skill practised (mathematics vs language) significantly predicted PES magnitude. The Tukey comparison test showed that greater PES was shown in learning activities practising mathematical skills ($M = 276$ ms, $SE = 12.4$ ms) than language skills ($M = 82.3$ ms, $SE = 11$ ms), $p = .02$ (Figure 1). The type of skill also moderated the PEA-PES association, $p < .001$. A separate linear mixed model for learning activities with mathematical skills showed no PES-PEA association, but for the learning activities practising language skills a higher PES average associated to greater proportion post-error correct, $\beta = 447.80$, $t(2011) = 5.59$, $p < .001$.

As an addition to this first model, we found that how much time participants were given to respond on each trial (time pressure) predicted PES, such that greater PES was found in learning activities with more time to respond, $\beta = 142.8$, $t(2942) = 8.27$, $p < .001$. Moreover, time to respond moderated the PES-PEA association positively, $\beta = 57.5$, $t(25027) = 13.36$, $p < .001$, meaning that with more time to respond participants showed more PES along with greater post-error accuracy (Figure 2B). The time to respond did not differ between the learning activities practising language ($M = 26.0$ s) and mathematical skills ($M = 29.2$ s), $p = .45$. Time pressure also did not interact with the type of skill practised, $p = .92$.

While all learning activities showed on average slower RT on error trials than on correct trials, there was an association between time to respond and error RT, such that longer time to respond associated with slower error RTs, $\beta = 0.46$, $t(1519124) = 422.8$, $p < .001$. Furthermore, when there was more time to respond, participants also exhibited slower errors compared to the RTs of the correct trials, $\beta = 0.15$, $t(1519124) = 189.1$, $p < .001$.

In addition, at the trial level, a model including the type of error (fast vs slow errors) was found to be a better model to explain PES than the basic model, $\chi^2 = 380.96$, $p < .001$. Here, participants showed greater PES effect after slow errors ($M = 276$ ms, $SE = 11.6$ ms) than after fast errors ($M = 147$ ms, $SE = 12.4$ ms), in line with what was found at the task level.

3.4 | Difficulty level (chosen by the participant)

At the participant level, we examined whether including the difficulty level chosen by the student (easy, medium, hard), which affects the selection of items and hence the error rate, would change the basic model (participant and learning activity as random effects) fit. There were no significant differences in age between the children choosing the difficulty levels. We found that adding difficulty level as a predictor significantly improved the model, $\chi^2 = 205.08$, $p < .001$. Figure 3 shows that participants choosing the most difficult level in learning activities also had the greatest PES effect ($M = 270$ ms, $SE = 12.6$ ms) compared to participants choosing the medium level ($M = 208$ ms, $SE = 11.8$ ms) and easy level ($M = 129$ ms, $SE = 10.8$ ms). Follow up comparisons using a Tukey test showed that participants choosing the hard level had a significantly greater PES than participants choosing the medium and the easy levels, p's $< .001$, and children choosing the medium level also showed significantly greater PES than children choosing the easy level, $p = .006$. To examine the PES-PEA association in comparison to the difficulty level, the expected proportion correct given the difficulty level was subtracted from the post-error correct metric. There was no

Developmental Science WILEY

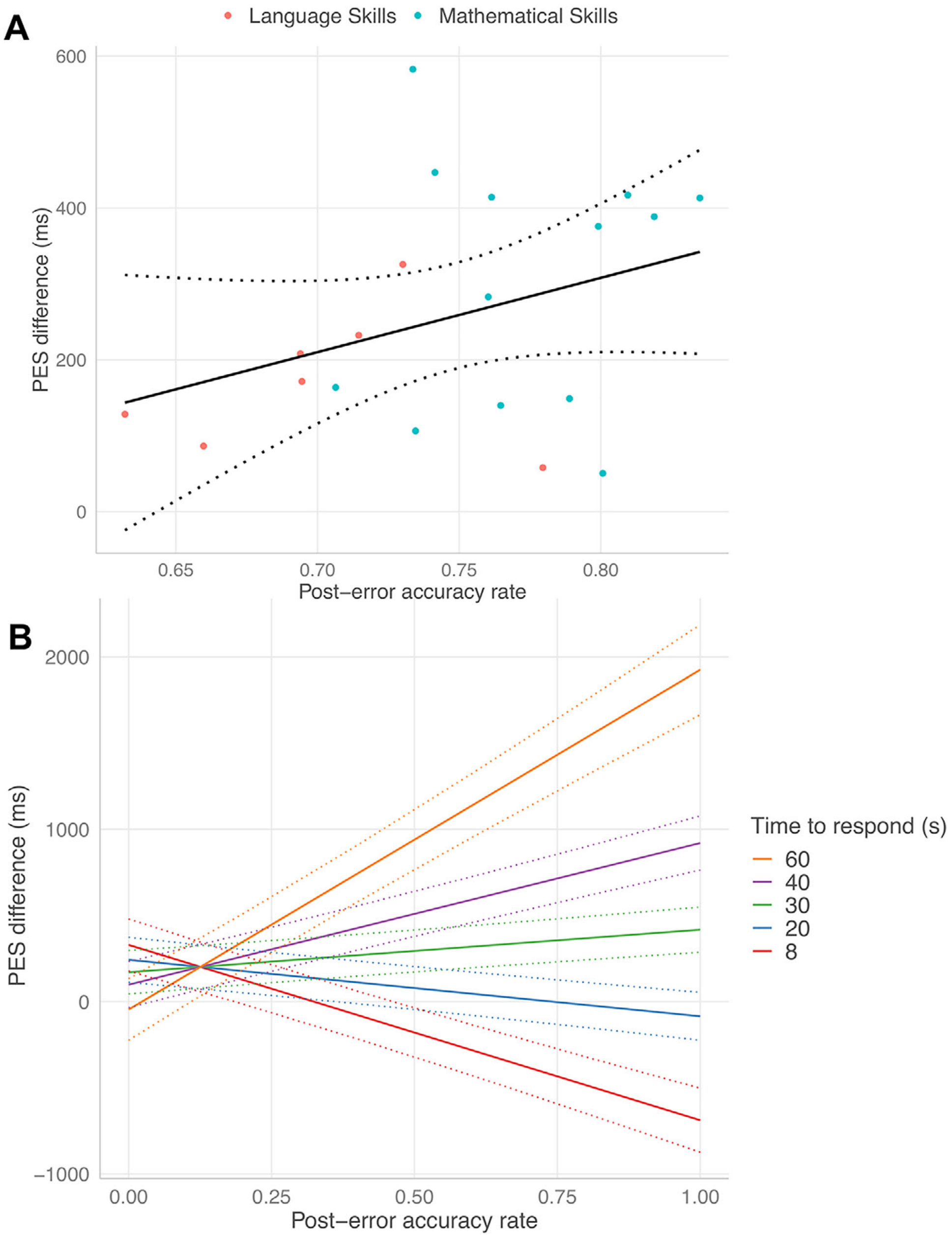

FIGURE 2 (A) Scatterplot of the association between mean PES_{diff} and mean proportion post-error accuracy across learning activities. (B) Association between post-error slowing (PES) difference (ms) and post-error accuracy as a function of how much time participants were given to respond in the learning activities. The lines represent the predicted PES-PEA associations for the different time limits. The dotted bands around the linear function represent the 95% confidence interval

interaction between the difficulty level and the PES-PEA association, such that even though participants showed more PES in the hard level, these participants did not show significantly more improved post-error performance in comparison to the participants in the other levels. But the PES-PEA association remained significant when controlling for difficulty level, $\beta = 131.4$, $t(122982) = 3.2$, $p = .001$.

3.5 | Age and ability

To investigate developmental differences in PES, we examined the association of PES magnitude with children's age and ability level at a participant level, above and beyond PEA. We found that age predicted PES linearly ($\beta = -50.6$, $p < .001$) and quadratically ($\beta = -26.5$, $p < .001$), where the magnitude of PES increased from 6 to 9 years old and decreased from 9 to 13 (Figure 4A). In the same model, the children's ability level was found to predict PES effect positively in a linear way ($\beta = 156.6$, $p < .001$), but not quadratically, $p = .60$ (Figure 4B). Neither the age nor the ability level of the child moderated the association between PES and PEA, age $p = .26$, ability level $p = .07$, but PEA remained a significant predictor of PES in the model, $\beta = 130.1$, $t(358880) = 3.15$, $p = .002$.

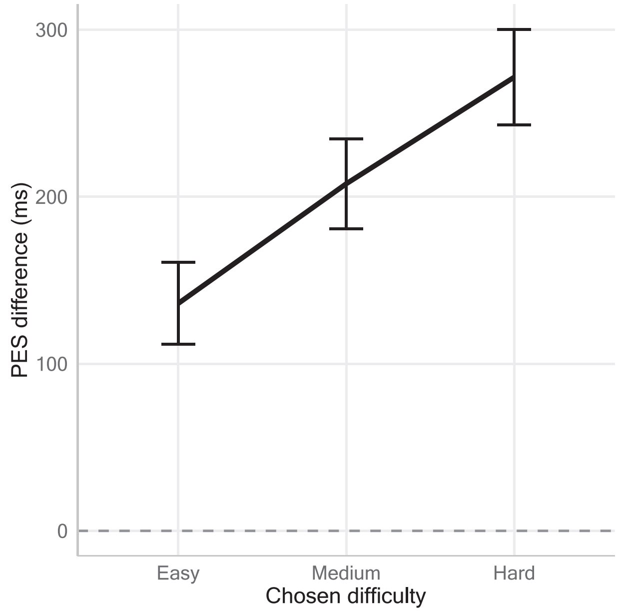

FIGURE 3 PES magnitude as a function of the difficulty level chosen by the participants. The difficult level corresponds to the probability that participants will give a correct answer: 90% at the easy level, 75% at the medium level and 60% at the hard level (60%). The vertical error bars represent the 95% confidence intervals

4 | DISCUSSION

In this study post-error slowing, which is thought to be a marker of adaptive behaviour, was investigated in an online learning environment for primary school children. In nearly all 23 mathematical and language-related learning activities we found post-error slowing (PES). To ensure a robust result, we used three different measures of PES, that consider the impact of fatigue, general response speed and the magnitude of the RT difference. We found no difference in the main results between these measures. PES associated with increased post-error accuracy (PEA), suggesting that children who slow down their performance after an error are adapting their strategies and improve their accuracy. Further analyses revealed variability in the magnitude of PES and its association with PEA as a function of task and individual differences. First, learning activities that involved practising mathematical skills showed the greatest PES effects, whereas during language skills practise the PES effect was smaller but had a greater impact on accuracy rate. Second, when children had more time to respond, that is, less time pressure, they also showed greater adaptive behaviour after an error. In line with this, we found that the PES effect was highest after making a slow error, in comparison to fast errors. Third, looking at individual differences between children we found that children who chose the most difficult practise level had the greatest PES, but not necessarily greater PEA than children in the other levels. Finally, we found that from 6 to 9 years of age, PES effects increased and from 9 to 13 years of age declined and that greater ability on the learning activities independently (in addition to the age effect) predicted greater PES magnitude.

Overall, the study showed promising findings for generalising PES in an educational context. The findings relating to the mathematically focused learning activities are in line with prior research examining PES within the context of mental arithmetic tasks (Borght et al., 2016; Desmet et al., 2012; Lavro et al., 2018b; Núñez-peña et al., 2017). This study supports the finding that children slow down their performance after an error in complex tasks in a range of mathematical tasks but more importantly, they also show this in an online practice system. No previous studies, to our knowledge, have addressed the PES effect in linguistic tasks. Although the effects were remarkably smaller, we also found PES in the linguistic domain. The substantial difference between the domains could be due to how this specific learning environment practises the particular skills. Further research is needed to support this finding using language-related tasks outside and inside the laboratory (e.g., spelling or reading).

Studying PES in an online learning environment has both advantages and limitations.

The advantages are that it is possible to examine PES in a large heterogeneous sample of children that play in a natural environment covering a wide range of tasks. Another advantage is the adaptivity of the practice system, which ensures that PES could be investigated in cognitively demanding tasks for all ages and ability levels. But there are also some drawbacks of collecting data in such a setting. Since children play in their natural environment, we cannot control or check their circumstances, such as the presence of any distraction in the room, device, or whether someone else is playing on their account. Since students can also practise the learning activity of their choice when they want, the data differ a lot in size between the learning activities and between the students. In this study, we controlled for these complex data by using learning activity and student as random effects in a linear mixed model.

Since we reliably observed post-error slowing across tasks, the next question was why it occurs and in which conditions. In this study, we show that students' accuracy was greater after an error when they showed greater post-error slowing, suggesting that PES may contribute to adaptive changes in behaviour after the detection of an error (Ullsperger et al., 2014), in line with the functional account proposed by the conflict monitoring theory (Botvinick et al., 2001; Dutilh et al., 2012a). This is in contrast with Notebaert's proposal that PES is a response to unexpected infrequent events, where the attention is shifted towards the errors, instead of towards improving performance (Notebaert et al., 2009). Moreover, this study showed that students performing at the most difficult level, with the highest probability of making an error, showed the greatest PES effect, which again is not in line with Notebaert's argument that PES diminishes as error frequency increases. It is debatable whether the impact of committing an error in this study is the same as for some of the previous studies. The orienting account was originally described for tasks where the errors represent impulsive incorrect response selection due to stimulus ambiguity, where errors are rare. Notably, in the tasks investigated here, the errors are more focused on learning, rather than performance, and problems become gradually more difficult for everyone instead of being simply repeated. Arguably, the type of error monitoring in a learning environment impacts the strategy use of students more. When

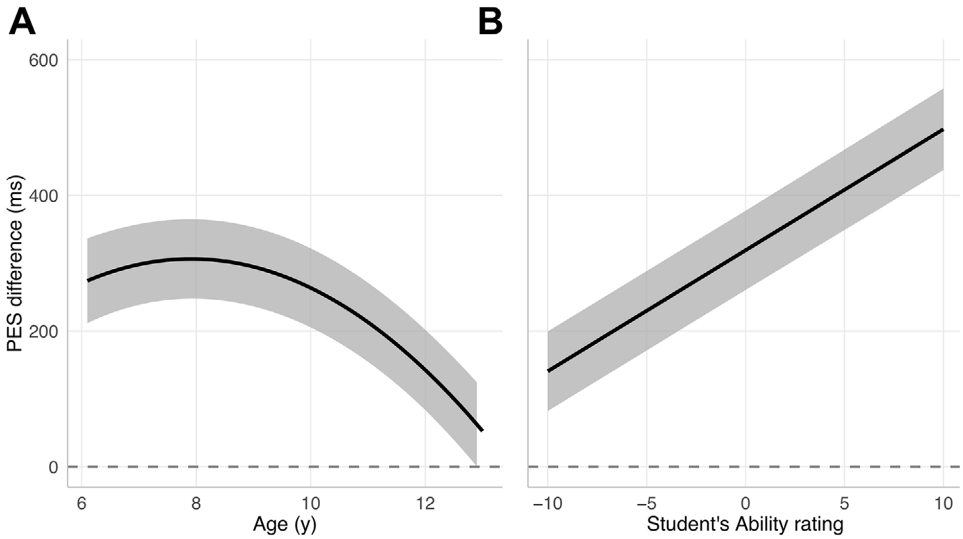

FIGURE 4 PES difference predicted by a combination of the linear and quadratic function of age (years) and ability rating. The grey bands around the functions represent the 95% confidence interval

children are learning, they typically use less sophisticated strategies first; after direct feedback (e.g., an error) they increase their reactive control, in line with the DMC framework, and try out more sophisticated strategies that might take longer to perform (Lemaire & Siegler, 1995), leading to greater PES. Children choosing the most difficult levels are more challenged in terms of learning new skills, that require more sophisticated strategies, which could account for the larger PES magnitude they show. This is also in line with our finding that students who show greater PES and may therefore be more closely monitoring their errors and adjusting their strategies, are the more able students who can solve more difficult problems. However, this latter part was not supported by a stronger PES-PEA relation for the most difficult level in comparison to the other levels.

In contrast to Damaso et al. (2020), we found that children show greater PES after slow errors than after fast errors. We also found that learning activities with longer time to respond and therefore less pressure to answer quickly and therefore less pressure to make fast errors showed bigger PES magnitude. This could again be in line with the use of more sophisticated strategies after an error, such that tasks with less pressure invite students to try out different strategies resulting in longer RTs and greater accuracy. What is different from previous studies such as Damaso et al. (2020), is that in this learning environment there is an emphasis on both speed and accuracy for all learning activities, therefore the influence of putting greater emphasis on speed, or accuracy, could not be investigated.

The age-related differences in PES showed that PES could be observed from the age of 6 until the investigated age of 13, in line with the study of Smulders et al. (2016). But contrary to this study, where they found stability with age, we found non-linear developmental differences, with increasing PES until the age of 9 followed by decline. Gupta et al. (2009) also found this developmental curve but with an earlier peak at the age of 7. However, between the age of 7 and 8 this decrease was not uniform, and the biggest reduction occurred between 9 and 10. Cognitive control continues to develop during childhood and children before the age of 8 are mostly relying on reactive control after an error, which is in line with the increasing PES that was observed in younger children in this study. From the age of 8 there is a developmental shift towards an improvement in and more use of proactive control skills, and away from reactive control, which may be reflected in the reduction of PES observed in late childhood in this study (Chevalier et al., 2015; Niebaum et al., 2020). This also goes jointly with a major development in executive functioning, especially task switching and error processing. Inhibition, which is required to withhold and delay a response, is thought to be the mechanism underlying error processing (Grammer, Carrasco, Gehring, & Morrison, 2014; Gupta et al., 2009; Marco-Pallarés et al., 2008). A major development in task switching takes place between 7 and 10 years of age, while major development in error processing occurs between 6 and 11 years of age. Before the age of 7, children are developing their ability to monitor and process errors accordingly, hence a large number of slow trials. After the age of 7, children are more able and faster, with less switch cost and inhibition, to recover from prior error trials causing a decrease in PES. For further research, it would be interesting to study the development of error processing longitudinally as well as its association with children's ability to switch between proactive and reactive control, and the development of their executive functioning, such as task switching and inhibitory control. More practically, given that PES is associated with improved performance, we could teach children to use more proactive strategies, especially in the context of a learning environment.

To conclude, we found an overall marker of adaptive behaviour, as measured by post-error slowing and post-error accuracy, in an online learning system. In light of the replication crisis (Open Science Collaboration, 2015) as well as the need for ecological validity in psychology (Lerner & Schmid Callina, 2014; Schmuckler, 2001), this study is valuable showing that the PES finding can be replicated and generalised to a variety of academic tasks. We also showed that the occurrence of this behaviour, depended on (1) the task, in terms of the amount of time pressure and skill practised; (2) previous trials practised, e.g., fast or slowly answered; (3) and characteristics of the learner, in terms of age, ability and motivation for challenge. This marker is a good proxy for whether children monitor their errors and adapt their behaviour and can be used to predict children's performance and progress in the skills practised.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 Marie Sklodowska-Curie Innovative Training Network (grant agreement number 721895).

CONFLICT OF INTEREST

Han L.J. van der Maas is full professor of Psychological Methods at the University of Amsterdam and founder of Oefenweb, the company that operates the online adaptive learning environment of Learn. Because he is not currently involved with this company and this study does not aim to demonstrate the effectiveness of training in the Learn system, we do not see a conflict of interest.

ETHICS

Schools and families with accounts are informed that Learn collects data for research purposes. Children (their parents or schools) can opt out of being part of the research done in the practice system and are, therefore, not included in this study. All data were anonymized before analysis. This procedure was approved by the Ethics Review Board of the department of Psychology at the University of Amsterdam.

DATA AVAILABILITY STATEMENT

Data and script for analysis are available on a dedicated research server of Prowise Learn in collaboration with University of Amsterdam. To get access, please email Susanne de Mooij: sdemoo01@mail.bbk.ac.uk .

ORCID

Susanne M. M. de Mooij https://orcid.org/0000-0001-8928-0851
Iroise Dumontheil https://orcid.org/0000-0001-5236-7849

REFERENCES

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1),. https://doi.org/10.18637/jss.v067.i01

Berwid, O. G., Halperin, J. M., Jonhson, R. J., & Marks, D. J. (2014). Preliminary evidence for reduced post-error reaction time slowing in hyperactive/inattentive preschool children. *Child Neuropsychology*, *20*(2), 196–209. https://doi.org/10.1080/09297049.2012.762760 PMID: 23387525

Van Der Borght, L., Desmet, C., & Notebaert, W. (2016). Strategy Changes After Errors Improve Performance. *Frontiers in Psychology*, *6*(2051), 1–6. https://doi.org/10.3389/fpsyg.2015.02051

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295x.108.3.624 PMID: 11488380

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010 PMID: 22245618

Brinkhuis, M. J. S., Savi, A. O., Hofman, A. D., Coomans, F., van der Maas, H. L. J., & Maris, G. (2018). Learning as it happens : A decade of analyzing and shaping a large-scale online learning system. *Journal of Learning Analytics*, *5*(2), 29–46. https://doi.org/10.18608/jla.2018.52.3

Ceccarini, F., & Castiello, U. (2018). The grasping side of post-error slowing. *Cognition*, *179*, 1–13. https://doi.org/10.1016/j.cognition.2018.05.026 PMID: 29886094

Chevalier, N., Martis, S. B., Curran, T., & Munakata, Y. (2015). Metacognitive processes in executive control development: The case of reactive and proactive control. *Journal of Cognitive Neuroscience*, *27*(6), 1125–1136. https://doi.org/10.1162/jocn PMID: 25603026

Chiew, K. S., & Braver, T. S. (2013). Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry. *Frontiers in Psychology*, *4*, 1–15. https://doi.org/10.3389/fpsyg.2013.00015 PMID: 23382719

Damaso, K., Williams, P., & Heathcote, A. (2020). Evidence for different types of errors being associated with different types of post-error changes. *Psychonomic Bulletin and Review*, 1–6. https://doi.org/10.3758/s13423-019-01675-w

Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. *Frontiers in Psychology*, *2*(233), 1–10. https://doi.org/10.3389/fpsyg.2011.00233 PMID: 21713130

de Mooij, S. M. M., Kirkham, N. Z., Raijmakers, M. E. J., van der Maas, H. L. J., & Dumontheil, I. (2020). Should online math learning environments be tailored to individuals' cognitive profiles? *Journal of Experimental Child Psychology*, *191*, 1–15. https://doi.org/10.1016/j.jecp.2019.104730

Denervaud, S., Knebel, J. F., Immordino-Yang, M. H., & Hagmann, P. (2020). Effects of traditional versus Montessori schooling on 4- to 15-year old children's performance monitoring. *Mind, Brain, and Education*, *14*(2), 167–175. https://doi.org/10.1111/mbe.12233

Desmet, C., Imbo, I., Brauwer, J. D.e, Brass, M., Fias, W., Notebaert, W., Desmet, C., Imbo, I., de Brauwer, J., Brass, M., & Fias, W. (2012). Error adaptation in mental arithmetic. *The Quarterly Journal of Experimental Psychology*, *65*(5), 1059–1067. https://doi.org/10.1080/17470218.2011.648943 PMID: 22439882

Dutilh, G., Ravenzwaaij, D. V., Nieuwenhuis, S., van der Maas, H. L. J., Forstmann, B. U., & Wagenmakers, E. (2012a). How to measure post-error slowing : A confound and a simple solution. *Journal of Mathematical Psychology*, *56*(3), 208–216. https://doi.org/10.1016/j.jmp.2012.04.001

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E. (2012b). Testing theories of post-error slowing. *Attention, Perception, and Psychophysics*, *74*, 454–465. https://doi.org/10.3758/s13414-011-0243-2

Fairweather, H. (1978). Choice reaction times in children : Error and post-error responses, and the repetition effect. *Journal of Experimental Child Psychology*, *26*, 407–418. https://doi.org/10.1016/0022-0965(78)90121-2

Ferdinand, N. K., & Kray, J. (2014). Developmental changes in performance monitoring : How electrophysiological data can enhance our understanding of error and feedback processing in childhood and adolescence. *Behavioural Brain Research*, *263*, 122–132. https://doi.org/10.1016/j.bbr.2014.01.029 PMID: 24487012

Gupta, R., Kar, B. R., & Srinivasan, N. (2009). Development of task switching and post-error-slowing in children. *Behavioral and Brain Functions*, *5*(1), 38. https://doi.org/10.1186/1744-9081-5-38 PMID: 19754947

Houtman, F., Castellar, E. N., & Notebaert, W. (2012). Orienting to errors with and without immediate feedback. *Journal of Cognitive Psychology*, *24*(3), 278–285. https://doi.org/10.1080/20445911.2011.617301

Jansen, B. R. J., Louwerse, J., Straatemeier, M., van der Ven, S. H. G., Klinkenberg, S., & van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, *24*, 190–197. https://doi.org/10.1016/j.lindif.2012.12.014

Jones, L. B., Rothbart, M. K., & Posner, M. I. (2003). Development of executive attention in preschool children. *Developmental Science*, *6*(5), 498–504. https://doi.org/10.1111/1467-7687.00307

Jonker, T. R., Seli, P., Cheyne, J. A., & Smilek, D. (2013). Performance reactivity in a continuous-performance task: Implications for understanding post-error behavior. *Consciousness and Cognition*, *22*(4), 1468–1476. https://doi.org/10.1016/j.concog.2013.10.005 PMID: 24177237

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects. *Journal of Statistical Software*, *82*(13),. https://doi.org/10.18637/jss.v082.i13 PMID: 29430216

Laming, D. (1979). Choice reaction performance following an error. *Acta Psychologica*, *43*, 199–224. https://doi.org/10.1016/0001-6918(79)90026-x

Lavro, D., Ben-Shachar, M. S., Saville, C. W. N., Klein, C., & Berger, A. (2018a). Testing the bottleneck account for post-error slowing beyond the post-error response. *Biological Psychology*, *138*, 81–90. https://doi.org/10.1016/j.biopsycho.2018.08.010 PMID: 30121286

Lavro, D., Levin, D., Klein, C., & Berger, A. (2018b). Response time distribution parameters show posterror behavioral adjustment in mental arithmetic. *Acta Psychologica*, *186*, 8–17. https://doi.org/10.1016/j.actpsy.2018.04.001 PMID: 29660604

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*(1), 83–97. https://doi.org/10.1037/0096-3445.124.1.83 PMID: 7897342

Lerner, R. M., & Schmid Callina, K. (2014). Relational developmental systems theories and the ecological validity of experimental designs: Commentary on freund and isaacowitz. *Human Development*, *56*(6), 372–380. https://doi.org/10.1159/000357179

Marco-Pallarés, J., Camara, E., Münte, T. F., & Rodríguez-Fornells, A. (2008). Neural mechanisms underlying adaptive actions after slips. *Journal of Cognitive Neuroscience*, *20*(9), 1595–1610. https://doi.org/10.1162/jocn.2008.20117 PMID: 18345985

Maris, G., & van der Maas, H. L. J. (2012). Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika*, *77*(4), 615–633. https://doi.org/10.1007/s11336-012-9288-y

Moser, J. S., Schroder, H. S., Heeter, C., Moran, T. P., & Lee, Y. (2011). Mind your errors: Evidence for a neural mechanism linking growth mind-set to adaptive posterror adjustments. *Psychological Science*, *22*(12), 484–489. https://doi.org/10.1177/0956797611419520

Niebaum, J. C., Chevalier, N., Guild, R. M., & Munakata, Y. (2020). Developing adaptive control: Age-related differences in task choices and awareness of proactive and reactive control demands. *Cognitive, Affective, and Behavioral Neuroscience*.

Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: An orienting account. *Cognition*, *111*(2), 275–279. https://doi.org/10.1016/j.cognition.2009.02.002 PMID: 19285310

Núñez-peña, M. I., Tubau, E., & Suárez-pellicioni, M. (2017). Post-error response inhibition in high math-anxious individuals: Evidence from a multi-digit addition task. *Acta Psychologica*, *177*, 17–22. https://doi.org/10.1016/j.actpsy.2017.04.002 PMID: 28431300

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 943–953. https://doi.org/10.1126/science.aac4716

Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *The Quarterly Journal of Experimental Psychology*, *29*(4), 727–743. https://doi.org/10.1080/14640747708400645

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356. https://doi.org/10.1111/1467-9280.00067

Regev, S., & Meiran, N. (2014). Post-error slowing is influenced by cognitive control demand. *Acta Psychologica*, *152*, 10–18. https://doi.org/10.1016/j.actpsy.2014.07.006 PMID: 25089881

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Cognition and Behavior*, *306*, 443–448. https://doi.org/10.1126/science.1100301

Schachar, R. J., Chen, S., Logan, G. D., Ornstein, T. J., Crosbie, J., Ickowicz, A., & Pakulak, A. (2004). Evidence for an error monitoring deficit in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, *32*(3), 285–293. https://doi.org/10.1023/b:jacp.0000026142.11217.f2 PMID: 15228177

Schmuckler, M. A. (2001). What is Ecological Validity? A Dimensional Analysis. *Infancy*, *2*(4), 419–436. https://doi.org/10.1207/S15327078IN0204_02 PMID: 33451194

Schroder, H. S., Nickels, S., Cardenas, E., Breiger, M., Perlo, S., & Pizzagalli, D. A. (2019). Optimizing assessments of post-error slowing: A neurobehavioral investigation of a flanker task. *Psychophysiology*, 1–17. https://doi.org/10.1111/psyp.13473

Schroder, H. S., Perlo, S., & Pizzagalli, D. A. (2020). Optimizing assessments of post - error slowing: A neurobehavioral investigation of a flanker task. *Psychophysiology*, *57*(2), 1–17. https://doi.org/10.1111/psyp.13473

Smulders, S. F. A., Soetens, E., & van der Molen, M. W. (2016). What happens when children encounter an error? *Brain and Cognition*, *104*, 34–47. https://doi.org/10.1016/j.bandc.2016.02.004 PMID: 26914174

Straatemeier, M. (2014). *Math Garden: A new educational and scientific instrument*. http://hdl.handle.net/11245/1.417091

Tamnes, C. K., Walhovd, K. B., Torstveit, M., Sells, V. T., & Fjell, A. M. (2013). Performance monitoring in children and adolescents: A review of developmental changes in the error-related negativity and brain maturation. *Developmental Cognitive Neuroscience*, *6*, 1–13. https://doi.org/10.1016/j.dcn.2013.05.001 PMID: 23777674

Team, R. C. (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. http://www.r-project.org/

Ullsperger, M., Danielmeier, C., & Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. *Physiological Reviews*, *94*(1), 35–79. https://doi.org/10.1152/physrev.00041.2012

van der Molen, M. W. (2000). Developmental changes in inhibitory processing: Evidence from psychophysiological measures. *Biological Psychology*, *54*, 207–239. https://doi.org/10.1016/s0301-0511(00)00057-0

Wilson, R. C., Shenhav, A., Straccia, M., & Cohen, J. D. (2019). The Eighty Five Percent Rule for optimal learning. *Nature Communications*, *10*(1), 1–9. https://doi.org/10.1038/s41467-019-12552-4

How to cite this article: de Mooij, S. M. M., Dumontheil, I., Kirkham, N. Z., Raijmakers, M. E. J., & van der Maas, H. L. J. (2021). Post-error slowing: Large scale study in an online learning environment for practising mathematics and language. *Developmental Science*, e13174. https://doi.org/10.1111/desc.13174

APPENDIX

POST ERROR SLOWING

A.1 Description of learning activities

	Name of the learning activity	Number of participants	Number of accuracy response sequences	Average age (y) (SD)	Description	Application	Response mode asked for question	Max time to response (s)
1	Addition	44949	104778	9.1 (1.9)	Sums from 1 + 1 = ... to 26900 + 4400 = ...	Mathematics	Mixed	20
2	Chaos of letters	13540	35278	9.4 (1.7)	Putting the letters of a word in the right order.	Language	Open answer	30
3	Counting	32395	67782	8.2 (1.9)	Counting the number of fishes on the screen [2-100 fishes].	Mathematics	Mixed	20
4	Dictation	11949	25077	9.8 (1.7)	Listening to a sentence, where a word is repeated afterwards. The repeated word needs to be written correctly.	Language	Open answer	30
5	Division	29109	78670	10.1 (1.5)	Divisions from 4 : 2 = ... to 4601 : 1000 = ...	Mathematics	Open answer	20
6	Flowercode	12912	32770	9.4 (1.7)	Logical reasoning game, similar to Mastermind, where the code needs to be cracked with limited but sufficient information.	Mathematics	Open answer	60
7	Fractions	4250	7136	10.9 (1.4)	Variety of exercises of fractions, percentages and proportions. For example, which of the five fractions is the smallest. Or, there is a 96% chance of rain today. What is the chance that it remains dry?	Mathematics	Mixed	30
8	Grammar	11892	31995	10.5 (1.4)	Naming the word class of a word in a sentence. For example, "I am giving the dog food. Is "the" a verb, article, noun or numeral?"	Language	Multiple Choice	20

(Continues)

	Name of the learning activity	Number of participants	Number of accuracy response sequences	Average age (y) (SD)	Description	Application	Response mode asked for question	Max time to response (s)
9	Money	46421	124599	9.6 (1.6)	Practise with coins and banknotes to estimate the price of a product. For example, what is the combined cost of a €1.50 ice cream and €2 fries?	Mathematics	Mixed	20/30/40
10	Multiplication	30789	65168	9.9 (1.5)	Multiplications from $1 \times 0 = \ldots$ to $6000 \times 803 = \ldots$..	Mathematics	Open answer	20
11	Numbers	7087	21292	9.6 (1.7)	Obtain a target number by using a set of provided numbers and operations. For example, obtain 1 by using the numbers 2, 4 and 9 and the operations × and – (solution is $2 \times 4 = 8$, $9-8 = 1$).	Mathematics	Open answer	60
12	Parsing words	9643	25549	10.9 (1.1)	Parsing the function of a word in the sentence. For example, "I will become a pilot". What is "I"? Direct/indirect object, the subject or the finite verb?	Language	Mixed	20
13	Proverbs	5007	12186	10.6 (1.6)	The right meaning of a proverb needs to be chosen out of five options.	Language	Multiple Choice	30
14	Reading	20547	70307	9.6 (1.7)	Reading a text and clicking on the nonsense words.	Language	Open answer	30
15	Series	9848	17490	9.0 (2)	Exercises where multiple operations are combined, e.g. $3 \times 5 + 2 = \ldots$ or very difficult, $(8-2) \times 10 : 5 \times 4$.	Mathematics	Mixed	20
16	Slow mix	29623	49447	9.2 (1.8)	Mix of arithmetical operations sums, subtractions, multiplications and divisions) at a slow pace.	Mathematics	Open answer	40

(Continues)

	Name of the learning activity	Number of participants	Number of accuracy response sequences	Average age (y) (SD)	Description	Application	Response mode asked for question	Max time to response (s)
17	Speed mix	18176	50898	10.1 (1.7)	Mix of arithmetical operations (sums, subtractions, multiplications and divisions) at a fast pace.	Mathematics	Multiple Choice	8
18	Spelling	27194	85865	10.0 (1.6)	Six different spellings of a word are presented. Five of them are spelled incorrectly and 1 is spelled correctly. The right spelling needs to be clicked.	Language	Multiple Choice	20
19	Subtraction	38590	82293	9.1 (1.9)	Subtractions from 8 – 8 = … to 85200 – 8870 = …..	Mathematics	Mixed	20
20	Telling time	23471	52975	9.4 (1.7)	Telling time with analogue and digital clocks.	Mathematics	Open answer	30
21	Verbs	14888	41235	10.7 (1.3)	Conjugating verbs in different tenses (present/past tense etc.). For example, "He [want]…an ice cream (present tense)."	Language	Open answer	30
22	Vocabulary	25718	74749	10.1 (1.5)	The right meaning of a word needs to be chosen out of five options. For example, courage = caring, bravery, cowardice, timid or honest.	Language	Multiple Choice	20
23	Word forms	18818	44609	10.3 (1.5)	Practicing setting words in the right singular/plural form, such as: "One belt. Five …."	Language	Open answer	30

A.2 Statistical results of the linear mixed models for PES_{diff} and $PES_{rel.}$

	PES_{diff}				PES_{rel}				PES_{robust}		
Learning activity	**B**	**SE**	***t***	***p***	**B**	**SE**	***t***	***p***	**Proportion *N* RTpost > RTpre**	χ^2	***p***
Addition	419.8	15.0	28.0	<0.001	0.05	0.002	28.4	<0.001	0.54	687.1	<0.001
Counting	63.2	13.8	4.6	<0.001	0.01	0.001	5.0	<0.001	0.51	34.6	<0.001
Dictation	190.3	29.8	6.4	<0.001	0.03	0.003	9.1	<0.001	0.52	39.7	<0.001
Division	418.0	17.4	24.0	<0.001	0.07	0.002	38.6	<0.001	0.55	943.4	<0.001
Flowercode	150.8	26.6	5.7	<0.001	0.02	0.003	5.7	<0.001	0.51	24.0	<0.001
Fractions	461.9	55.6	8.3	<0.001	0.04	0.006	7.4	<0.001	0.53	33.0	<0.001
Grammar	**43.5**	**26.8**	**1.6**	**0.10**	0.01	0.003	2.7	0.008	**0.50**	**2.2**	**0.075**
LetterChaos	**−41.1**	**25.5**	**−1.6**	**0.11**	**−0.01**	**0.002**	**2.4**	**0.02**	**0.50**	**2.2**	**0.930**
Money	299.7	9.0	33.1	<0.001	0.03	0.001	27.5	<0.001	0.54	450.9	<0.001
Multiplication	385.6	18.7	20.6	<0.001	0.05	0.002	25.4	<0.001	0.54	451.0	<0.001
Numbers	115.3	33.2	3.4	0.001	0.01	0.004	2.6	0.01	0.51	4.0	0.022
Parsing words	81.5	29.8	2.7	0.006	0.01	0.003	3.3	<0.001	0.51	4.4	0.018
Proverbs	109.7	42.8	2.6	0.01	0.01	0.005	2.3	0.04	0.51	7.8	0.003
Reading	68.5	18.8	3.6	<0.001	0.01	0.002	4.8	<0.001	0.51	23.5	<0.001
Series	428.1	35.4	12.1	<0.001	0.05	0.004	12.9	<0.001	0.55	163.5	<0.001
Slow mix	6119	21.2	28.9	<0.001	0.06	0.002	26.2	<0.001	0.55	403.4	<0.001
Speed mix	154.9	21.7	7.2	<0.001	0.04	0.002	16.8	<0.001	0.54	250.5	<0.001
Spelling	**−87.4**	**16.9**	**−5.2**	**<0.001**	**−0.01**	**0.002**	**8.0**	**<0.001**	**0.49**	**49.3**	**1.000**
Subtraction	438.6	16.8	26.2	<0.001	0.05	0.002	29.5	<0.001	0.54	628.9	<0.001
Telling time	150.4	20.7	7.3	<0.001	0.01	0.002	5.6	<0.001	0.51	16.9	<0.001
Verbs	232.6	23.7	9.8	<0.001	0.03	0.003	10.9	<0.001	0.52	76.7	<0.001
Vocabulary	175.4	18.0	9.8	<0.001	0.02	0.002	11.8	<0.001	0.52	92.4	<0.001
Word forms	320.5	22.6	14.2	<0.001	0.05	0.002	18.6	<0.001	0.53	210.0	<0.001

Separate proportion tests were performed for PES_{robust} (prop > 0.50). Highlighted in bold are the learning activities that are not significantly showing PES.